```
               GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:       July 18, 2001, 00:04:02 ; Search time 1176.35 Seconds
                                        (without alignments)
                                        2892.870 Million cell updates/sec

Title:           US-09-457-066-1_COPY_829_1188
Perfect score:   360
Sequence:        1 gctttgtttttggaagaat.........gcagagggagcacaggagga 360

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        1028115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters:      20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
```

1: gb_est1:*		44: em_esthum10:*
2: gb_est2:*		45: em_esthum11:*
3: gb_est3:*		46: em_esthum12:*
4: gb_est4:*		47: em_esthum13:*
5: gb_est5:*		48: em_esthum14:*
6: gb_est6:*		49: em_esthum15:*
7: gb_est7:*		50: em_esthum16:*
8: gb_est8:*		51: em_esthum17:*
9: gb_est9:*		52: em_esthum18:*
10: gb_est10:*		53: em_esthum19:*
11: gb_est11:*		54: em_esthum20:*
12: gb_est12:*		55: em_esthum21:*
13: gb_est13:*		56: em_esthum22:*
14: gb_est14:*		57: em_esthum23:*
15: gb_est15:*		58: em_esthum24:*
16: gb_est16:*		59: em_esthum25:*
17: gb_est17:*		60: em_esthum26:*
18: gb_est18:*		61: em_esthum27:*
19: gb_est19:*		62: em_esthum28:*
20: gb_est20:*		63: em_estin1:*
21: gb_est21:*		64: em_estin2:*
22: gb_est22:*		65: em_estin3:*
23: gb_est23:*		66: em_estin4:*
24: gb_est24:*		67: em_estin5:*
25: gb_est33:*		68: em_estom1:*
26: gb_est34:*		69: em_estom2:*
27: gb_est35:*		70: em_estov1:*
28: gb_est36:*		71: em_estov2:*
29: gb_est37:*		72: em_estpl1:*
30: gb_est38:*		73: em_estpl2:*
31: gb_est39:*		74: em_estpl3:*
32: gb_est40:*		75: em_estpl4:*
33: em_estba:*		76: em_estpl5:*
34: em_estfun:*		77: em_estpl6:*
35: em_esthum1:*		78: em_estpl7:*
36: em_esthum2:*		79: em_estpl8:*
37: em_esthum3:*		80: em_estpl9:*
38: em_esthum4:*		81: em_estpl10:*
39: em_esthum5:*		82: em_estro1:*
40: em_esthum6:*		83: em_estro2:*
41: em_esthum7:*		84: em_estro3:*
42: em_esthum8:*		85: em_estro4:*
43: em_esthum9:*		86: em_estro5:*
		87: em_estro6:*
		88: em_estro7:*
		89: em_estro8:*
		90: em_estro9:*
		91: em_estro10:*
		92: em_estro11:*
		93: em_estro12:*
		94: em_estro13:*
		95: em_estro14:*
		96: em_estro15:*
		97: em_estro16:*
		98: em_estro17:*
		99: em_estro18:*
		100: em_estro19:*
		101: em_estro20:*
		102: gb_est25:*
		103: gb_est26:*
		104: gb_est27:*
		105: gb_est28:*
		106: gb_est29:*
		107: gb_est30:*
		108: gb_est31:*
		109: gb_est32:*
		110: gb_est41:*
		111: gb_est42:*
		112: gb_est43:*
		113: gb_est44:*
		114: gb_est45:*
		115: gb_est46:*
		116: gb_est47:*

```
117:  gb_est48:*          190:  gb_est110:*
118:  gb_est49:*          191:  gb_est111:*
119:  gb_est50:*          192:  gb_htc:*
120:  gb_est51:*          193:  em_gss_fun:*
121:  gb_est52:*          194:  em_gss_hum1:*
122:  gb_est53:*          195:  em_gss_hum2:*
123:  gb_est54:*          196:  em_gss_hum3:*
124:  gb_est55:*          197:  em_gss_hum4:*
125:  gb_est56:*          198:  em_gss_hum5:*
126:  gb_est57:*          199:  em_gss_hum6:*
127:  gb_est58:*          200:  em_gss_hum7:*
128:  gb_est59:*          201:  em_gss_hum8:*
129:  gb_est60:*          202:  em_gss_hum9:*
130:  gb_est61:*          203:  em_gss_inv1:*
131:  gb_est62:*          204:  em_gss_inv2:*
132:  gb_est63:*          205:  em_gss_inv3:*
133:  gb_est64:*          206:  em_gss_other:*
134:  gb_est65:*          207:  em_gss_pln1:*
135:  gb_est66:*          208:  em_gss_pln2:*
136:  gb_est75:*          209:  em_gss_pro:*
137:  gb_est76:*          210:  em_gss_rod1:*
138:  gb_est77:*          211:  em_gss_rod2:*
139:  gb_est78:*          212:  em_gss_rod3:*
140:  gb_est79:*          213:  em_gss_rod4:*
141:  gb_est80:*          214:  em_gss_rod5:*
142:  gb_est81:*          215:  em_gss_vrt1:*
143:  gb_est82:*          216:  em_gss_vrt2:*
144:  gb_est83:*          217:  em_gss_vrt3:*
145:  gb_est84:*          218:  gb_gss1:*
146:  gb_est85:*          219:  gb_gss2:*
147:  gb_est86:*          220:  gb_gss3:*
148:  gb_est87:*          221:  gb_gss4:*
149:  gb_est88:*          222:  gb_gss5:*
150:  gb_est89:*          223:  gb_gss6:*
151:  gb_est90:*          224:  gb_gss7:*
152:  gb_est99:*          225:  gb_gss8:*
153:  gb_est100:*         226:  gb_gss9:*
154:  gb_est101:*         227:  gb_gss10:*
155:  gb_est102:*         228:  gb_gss11:*
156:  gb_est103:*         229:  gb_gss12:*
157:  gb_est104:*         230:  gb_gss13:*
158:  gb_est105:*         231:  gb_gss14:*
159:  gb_est106:*         232:  gb_gss15:*
160:  gb_est67:*          233:  gb_gss16:*
161:  gb_est68:*          234:  gb_gss17:*
162:  gb_est69:*          235:  gb_gss18:*
163:  gb_est70:*          236:  gb_gss19:*
164:  gb_est71:*          237:  gb_gss20:*
165:  gb_est72:*          238:  gb_gss21:*
166:  gb_est73:*          239:  gb_gss22:*
167:  gb_est74:*          240:  gb_gss23:*
168:  gb_est91:*          241:  gb_gss24:*
169:  gb_est92:*          242:  gb_gss25:*
170:  gb_est93:*          243:  gb_gss26:*
171:  gb_est94:*          244:  gb_gss27:*
172:  gb_est95:*          245:  gb_gss28:*
173:  gb_est96:*          246:  gb_gss29:*
174:  gb_est97:*          247:  gb_gss30:*
175:  gb_est98:*          248:  gb_gss31:*
176:  em_esthum29:*       249:  gb_gss32:*
177:  em_esthum30:*       250:  gb_gss33:*
178:  em_esthum31:*       251:  gb_gss34:*
179:  em_esthum32:*       252:  em_gss_inv4:*
180:  em_esthum33:*       253:  em_gss_rod6:*
181:  em_estom3:*         254:  em_gss_rod7:*
182:  em_estpl11:*        255:  em_gss_rod8:*
183:  em_estro21:*        256:  gb_gss35:*
184:  em_estro22:*        257:  gb_gss36:*
185:  em_estro23:*        258:  gb_gss37:*
186:  em_htc:*
187:  gb_est107:*
188:  gb_est108:*
189:  gb_est109:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
                                    Email: cgapbs-r@mail.nih.gov
                                    Tissue Procurement: ATCC
                                    cDNA Library Preparation: CLONETECH Laboratories Inc.
                                    cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                                    DNA Sequencing by: Incyte Genomics, Inc.
                                    Clone distribution: MGC clone distribution information can be
                                    found through the I.M.A.G.E. Consortium/LLNL at:
                                    http://image.llnl.gov
                                    Plate: LLCM763  row: m  column: 22
                                    High quality sequence start: 2
                                    High quality sequence stop: 513.
FEATURES             Location/Qualifiers
     source          1..523
                                    /organism="Homo sapiens"
                                    /db_xref="taxon:9606"
                                    /clone="IMAGE:3930045"
                                    /clone_lib="NIH_MGC_56"
                                    /tissue_type="primitive neuroectoderm"
                                    /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                                    SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattggcc).
                                    Double-stranded cDNA was prepared from cell line RNA. 5'
                                    and 3' adaptors were used in cloning as follows: 5'
                                    adaptor sequence: 5'-CACGGCCATTATGGCC-3' adaptor
                                    sequence: 5'-ATTCTAGAGGCGGCCGGCCGACATG-dT(30)BN-3'
                                    (where B = A, C, or G and N = A, C, G, or T). Average
                                    insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
                                    contained inserts by PCR. This library was enriched for
                                    full-length clones and was constructed by Clontech
                                    Laboratories (Palo Alto, CA)."
BASE COUNT     147 a      118 c      129 g      128 t      1 others
ORIGIN

     Query Match                        95.5%;  Score 343.8;  DB 142;  Length 523;
     Best Local Similarity  98.9%;  Pred. No. 8.1e-97;
     Matches  356;  Conservative  0;  Mismatches  3;  Indels  1;  Gaps  1;

Qy      1  gctttgtttttggaagaaatccagagtggtgatctgaacctcaacagaggagta     60
           |||||||||||||  |||||||||||||||||||||||||| ||||| ||||||
Db    151  GCTTCTGTTTTCGGAAGAAAATCCAGAGTGGTGATCTGAACCTCTAACAGAGGAGTA    210

Qy     61  agattatacagctgcacacctcgtaacttctcagtgtccataaggagaactaaagaga    120
           ||||||||||||| |||||||||||||||||||||||||||| |||||||| |||||
Db    211  AGATTATACAGCTGCACACCTCGTAACTTCTCAGTGTCCATAAGGAGAAGAACTAAAGAGA    270

Qy    121  acggtacacatttctggccaggttgtctcctggtaaacgctggtggtgggaactgtgcc    180
           |||| |||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    271  ACCGATACCATTTTCTGGCCCAGGTGTGTCTCCTGGTTAAACCGTGTTGGTGGGAACTGTGCC    330

Qy    181  tgtgtctccacaattgcaatgatgatgcaaagaccggtcagggggattgcacaaatcacctcacc    240
           || ||||||||||||||||||||||| |||||||||||||||||||||||||||| ||||||
Db    331  TGNTGTCTCCACAATTGCAATGATGAATGTCAATGTCCCAAGCAAAGTTACTTACTTACTAAAAATAC    390

Qy    241  cacgaggtccttcagttgagacccaaagaccggtcaggggattgcacaaatcacctcacc    300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    391  CACGAGGTCCTTCAGTTGAGACCCAAAGACCGGTCAGGGGATTCAGGGATTCACAAAATCACTCACC    450

Qy    301  gacgtggccctgaacgcaccatgagagtgtactgtgtgtgcagagggagcacaggagga    360
           |||||||  |||||||||||||||||||||||||||||||||||||||||||||||||
Db    451  GACGTGG-CCTGGAGCACCATGAGAGTGTGACTGTGTGTGTGCAGAGGGAGCACAGGAGGA    509

RESULT  2
AL047637
LOCUS       AL047637           304 bp    mRNA              EST       01-MAR-2000
DEFINITION  DKFZp586J0421_r1 586 (synonym: hutel) Homo sapiens cDNA clone
            DKFZp586J0421, mRNA sequence.
ACCESSION   AL047637
VERSION     AL047637.1  GI:4728633
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
```

SUMMARIES

Result NO.		Score	Query % Match	Length	DB	ID	Description
1		343.8	95.5	523	142	BE958470	BE958470 601644787
2		302.4	84.0	304	105	AL047637	AL047637 DKFZp586J
3		302	83.9	851	144	BF102859	BF102859 601646827
4		283.8	78.8	950	143	BF031624	BF031624 601558104
5		274.2	76.2	523	143	BF021679	BF021679 uy50f05.y
6		265.2	73.7	556	155	BG609411	BG609411 323251 MA
7		261.8	72.7	564	145	BF151355	BF151355 uz15b12.y
8		255.8	71.1	874	168	BF670092	BF670092 602119417
9		250.2	69.5	910	175	BG243001	BG243001 602355974
10		230.2	63.9	447	143	BF011835	BF011835 us37d10.y
11		209.2	58.1	968	174	BG118707	BG118707 602348280
12		158.4	44.0	902	145	BF163629	BF163629 601769732
13		141.4	39.3	425	11	AA759138	AA759138 ah78f03.s
14		139.6	38.8	521	166	BE374398	BE374398 601227568
c 15		118.6	32.9	324	15	AI020581	AI020581 ua96c08.r
c 16		116.6	32.4	877	220	CNS024MX	AL180978 Tetraodon
17		116.4	32.3	337	112	AW167693	AW167693 xn47g09.x
18		116.4	32.3	389	15	AI024617	AI024617 ov52f10.x
19		96.4	26.8	511	113	AW210331	AW210331 ul53f09.y
20		94.6	26.3	126	144	BF117246	BF117246 uz07a06.y
21		89.4	24.8	354	20	AI470494	AI470494 tj44e11.x
22		89	24.7	1796	192	AK003359	AK003359 Mus muscu
23		83.8	23.3	192	118	AW582560	AW582560 RC1-ST027
c 24		75.6	21.0	1036	221	CNS04S8A	AL304867 Tetraodon
25		69.4	19.3	765	145	BF143048	BF143048 601788733
26		68.6	19.1	155	142	BE937452	BE937452 RC1-ST027
27		54.2	15.1	649	165	BE289243	BE289243 601092806
28		53	14.7	560	122	AW968905	AW968905 EST380981
29		41	11.4	412	190	W21436	W21436 zb52a06.r1
30		39.6	11.0	1029	174	BG173847	BG173847 602333906
c 31		38.8	10.8	828	221	CNS03ULN	AL261284 Tetraodon
32		36.6	10.2	877	145	BF137533	BF137533 601780532
33		36	10.0	696	153	BG430400	BG430400 602502113
34		34.4	9.6	584	240	AZ243830	AZ243830 RPCI-23-4
c 35		34.2	9.5	444	143	BE990576	BE990576 UI-M-B1-
c 36		34.2	9.5	445	143	BE990947	BE990947 UI-M-B1-
37		34.2	9.5	547	141	BE844607	BE844607 EST352 Ap
38		34.2	9.5	1059	221	CNS03Y1Z	AL265760 Tetraodon
c 39		33.8	9.4	400	241	AZ279383	AZ279383 RPCI-23-1
c 40		33.8	9.4	479	190	W76903	W76903 me58b06.r1
41		33.8	9.4	506	256	B78465	B78465 T30C19TFB T
c 42		33.6	9.3	149	142	BE937398	BE937398 RC1-ST027
c 43		33.6	9.3	336	108	AV024535	AV024535 AV024535
44		33.4	9.3	357	7	AA424616	AA424616 zv91c03.r
45		33.4	9.3	481	148	BF433302	BF433302 7q60f09.x

ALIGNMENTS

```
RESULT  1
BE958470
LOCUS       BE958470           523 bp    mRNA              EST       04-OCT-2000
DEFINITION  601644787F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:3930045 5',
            mRNA sequence.
ACCESSION   BE958470
VERSION     BE958470.1  GI:10569175
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 523)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
REFERENCE   1 (bases 1 to 851)
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM921  row: f   column: 08
            High quality sequence stop: 529.
FEATURES    Location/Qualifiers
     source 1..851
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4073095"
            /clone_lib="NIH_MGC_60"
            /tissue_type="adenocarcinoma"
            /lab_host="DH10B (T1 phage-resistant)"
            /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
            Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
            ); Double-stranded cDNA was prepared from cell line RNA.
            5' and 3' adaptors were used in cloning as follows: 5'
            adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
            sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
            (where B = A, C, or G and N = A, C, G, or T). Average
            insert size 1.5 kb (range 0.9-4.0 kb). 1/15 colonies
            contained inserts by PCR. This library was enriched for
            full-length clones and was constructed by Clontech
            Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
            Library."
BASE COUNT    234 a    169 c    238 g    210 t
ORIGIN

Query Match            83.9%;  Score 302;  DB 144;  Length 851;
Best Local Similarity  96.1%;  Pred. No. 1.1e-83;
Matches  342;  Conservative   0;  Mismatches  10;  Indels  4;  Gaps  3;

Qy    5 ttgtttttggagaaatccagagtggtggatctgaaccttcaacgagggggtaagat 64
         || ||||||||||||||||||||||| |||||||||||||||||||| |||||
Db  199 TTTGTTTTGGAGAAGAAATCCAGAGTGGTGGATCTGAACCTTCAACAGAGGAGGTAAGAT 258

Qy   65 tatacagctgcacacctgtaacttcagtgtccataagggaagaactaaagagaaccg 124
         |||||||||||||| ||||||||| |||||||||||||||||| ||||||||||
Db  259 TATACAGCTGCACACCTCGTAACTTCCAGTGTCCATAAGGGAAGAACTAAAGAGAACCG 318

Qy  125 ataccatttctggccaggtgtctcctggttaaacgctgtgtgtgggaacttgcctgtt 184
         ||||||| ||||| ||||| |||||||||||||||||||||||||||||||| |
Db  319 ATACCATTTCTCGGCCCAGGTTGCTCCTGGTTAAACGCTGTGTGTGGGAACTTGCCTG-T 377

Qy  185 gtctccacaattgcaatgaatgtcaaagtgtcccaagcaaagttactaaaaatacacg 244
         |||||||||||||||||||||| ||||||||||||||||||| |||||||||||
Db  378 GTCTCCACAATTGCAATGAATGTCAATGTGTCCCAAGCAAAGTTACTAAAAATACCACG 437

Qy  245 aggtccttcagttgagaccaaagaccggtcagggggattgcacaaatcaccaccgacg 304
         |||||||||||||||||||||||||| ||||||| |||||||||||||| ||||||||
Db  438 AGGTCCTTCAGTTGAGACCAAAGACCGGTGTCAGGGGA-TGCACAAATCAC-CACCGACG 496

Qy  305 tggccctggagcacccatgagggagtgtgactgtgtgcagagggagcacaggagga 360
         |||||   ||||||||||| ||||||||||||||||| |||||||| ||||| |||
Db  497 TGGCC--TGAGCACCATGAGGAGTGTGACTGTGTGCGAGGGGAGCCACAGGGGGA 550

RESULT 4
BF031624
LOCUS       BF031624   950 bp   mRNA        EST       10-OCT-2000
DEFINITION  601158104F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3827760 5',
```

```
REFERENCE   1 (bases 1 to 304)
AUTHORS     Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
TITLE       EST (Ottenwaelder, et al.)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Ottenwaelder B
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. No s1 sequence
            available.
            This clone (DKFZp586J0421) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES    Location/Qualifiers
     source 1..304
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="DKFZp586J0421"
            /clone_lib="586 (synonym: hutel)"
            /tissue_type="uterus"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Vector: pSport1; Site_1: NotI; Site_2: SalI/MluI"
BASE COUNT     86 a     68 c     80 g     70 t
ORIGIN

Query Match            84.0%;  Score 302.4;  DB 105;  Length 304;
Best Local Similarity  99.7%;  Pred. No. 6.5e-84;
Matches  303;  Conservative   0;  Mismatches   0;  Indels  1;  Gaps  0;

Qy   51 agaggaggtaagattatacagctgcacacctgtaacttctcagtgtccataagggaga 110
         ||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    1 AGAGGAGGTAAGATTATACAGCTGCACACCTCGTAACTTCTCAGTGTCCATAAGGGAAGA 60

Qy  111 actaaagagaaccgatacctttctggccaggttgtctcctggtggtgg 170
         |||||||||||||||||| |||||||||||||||||||||||  ||||
Db   61 ACTAAAGAGAACCGATACCATTTCTGGCCAGGTTGTCTCCTGGTTAAACGCTGTGGTGG 120

Qy  171 gaactgtgctgttgtctctccacaattgcaatgatgtcccaagcaaagtta 230
         |||||||| ||||||||||||||||||||||||||||||||||||||||
Db  121 GAACTGTGCCTGTTGTCTCTCCACAATTGCAATGATGTCAATGTGTCCCAAGCAAAGTTAC 180

Qy  231 taaaaatacaccacgaggtcttcagtgtcctgagaccaccggtgtgcacaa 290
         |||||||||||||||||||| |||||||| ||||||||||||||||||||
Db  181 TAAAAATACCACCGAGGTCCTTCAGTTGTCTCCAGTCCAAGCAAAGGGGATTGCACAA 240

Qy  291 atcactcaccgacgtgccctggagcaccataggagtgtgactgtgtgcagagggag 350
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ATCACTCACCGACGTGCCCTGGAGCACCATGAGGAGTGTGACTGTGTGCAGAGGGAG 300

Qy  351 caca 354
         ||
Db  301 CTCA 304

RESULT 3
BF102859
LOCUS       BF102859   851 bp   mRNA        EST       19-OCT-2000
DEFINITION  601646827F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4073095 5',
            mRNA sequence.
ACCESSION   BF102859
VERSION     BF102859.1  GI:10885385
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Wed Jul 18 08:51:37 2001   us-09-457-066-1_copy_829_1188.rst

LOCUS   BF021679   523 bp   mRNA   EST 29-DEC-2000
DEFINITION   uy50f05.y1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:3663009 5'
               similar to TR:Q9QY71 Q9QY71 FALLOTEIN. ; mRNA sequence.
ACCESSION   BF021679
VERSION   BF021679.1 GI:10753011
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 523)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
               Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Gilbert Smith, Ph.D.
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Washington University Genome Sequencing Center
               Clone distribution: NCI-CGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               image.llnl.gov/image/html/iresources.shtml
               MGI:1423777
               Seq primer: -40RP from Gibco
               High quality sequence stop: 452.
FEATURES   Location/Qualifiers
     source   1..523
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3663009"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
                     dT. Library constructed by Life Technologies, Inc."
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT   143 a   116 c   139 g   125 t
ORIGIN

Query Match   76.2%;   Score 274.2; DB 143; Length 523;
Best Local Similarity   85.2%;   Pred. No. 5e-75;
Matches 306; Conservative 0; Mismatches 53; Indels 0; Gaps 0; .

Qy   1 gcttttgttttggagaaatccagagtggtgatctgaacctctaacagaggagta 60
          ||||| |||||| ||||||||| |||||| |||| || |||||| || |||| |||
Db   164 GCTTTCCTGTATGGGAAAAAAGCAAGTGTGAATCTGAATCTCCTAAAGGAAGAGGTA 223

Qy   61 agattatacagctgcacacctgtaacttccagtgtccataagggaagaactaaagaga 120
          || || |||||||||||||||||| ||||| |||||||| ||||||||| |||||||||
Db   224 AAACTCTACAGCTGCACACCTGCACACCCCGGAACTTCCAGTGTCCATACGGGAAGAGG 283

Qy   121 accgataccatttctggccaggttgtctccggttaaacgctggtggtgggaactgtgcc 180
          ||||| ||||| ||||||||||||||||||||| ||||||||||||||||| ||| |||
Db   284 ACAGATACCATTGTCTGGCCAGGTTGTCTCCTGGTCAAGCCGGTGCAAGCCGTGTGCC 343

Qy   181 tgtgtctccacaattgcaatgatgtcaatgtcccaagcaaagttactaaaaatac 240
          ||||| ||||||||||| |||||||| ||| ||||| |||||| || ||| ||||||
Db   344 TGTTGTCTCCATAATTGCAATGAATGTCAGTGTGTCCCACCTAAAGTTACAAAAAGTAC 403

Qy   241 cacgggtccttcagttgagaccaaagaccgtgtcagggggattgcacaaatcactcacc 300
          || || ||||||||| ||||||||| |||||||| ||||||| ||||| ||| ||||||
Db   404 CATGAGGTCCTTCAGTTGACCTAGACCAAAAACTGGAGTCAAGGGGATTGCATAAGTCACTCACT 463

Qy   301 gacgtggccctggacgccatgaggagtgtgactgtgtgcagagggagcacaggagg 359
          || ||||||||| |||||| ||||| |||||| |||||||||||||||||| |||||||
Db   '464 GATGTGGCCTCTGGAACACCCAGGAGAATGTGACTGGTGTGTGTAGAGGAAACCCAGGAGG 522

RESULT 5
BF021679

```
RESULT         6
BG609411
LOCUS          BG609411       556 bp   mRNA                    EST        17-APR-2001
DEFINITION     323251 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION      BG609411
VERSION        BG609411.1  GI:13659390
KEYWORDS       EST.
SOURCE         pig.
ORGANISM       Sus scrofa
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE      1  (bases 1 to 556)
AUTHORS        Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
               Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
               and Keele,J.W.
TITLE          Design and use of two pooled tissue normalized cDNA libraries for
               EST discovery in swine
JOURNAL        Unpublished (2000)
COMMENT        Contact: Smith TPL
               USDA, ARS, US Meat Animal Research Center
               PO Box 166, Clay Center, NE 68933-0166, USA
               Tel: 402 762 4366
               Fax: 402 762 4390
               Email: smith@email.marc.usda.gov
               Single pass sequencing. Bases called and alt_trimmed with phred
               v0.980904.e. Vector identified by cross_match with the -minscore 18
               and -minmatch 12 options.
               PCR Primers
               FORWARD: AGGAAACAGCTATGACCAT
               BACKWARD: GTTTTCCCAGTCACGACG
               Plate: 98  row: P  column: 8
               Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..556
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
BASE COUNT     139 a    153 c    141 g    123 t
ORIGIN

Query Match             73.7%;  Score 265.2;  DB 155;  Length 556;
Best Local Similarity   86.0%;  Pred. No. 3.4e-72;
Matches  294;  Conservative   0;  Mismatches  48;  Indels   0;  Gaps    0;

Qy      1 gctttgttttggaagaaaatccagagtggtggatctgaacctctaacagagaggagta 60
          ||||| |||||||||||| || ||||||||| || |||||| || || |  |||||||||
Db    215 GCTTTTGTTTTCGGAGAAAATCCAGAGTGGTGGATCGAATCGAACCTTCTCAAAGAAGAAGTG 274

Qy     61 agattacagctgcacacctgtaacctctcagtgtccataaggaagaactaaagaga 120
          || |||||||||| ||| |||| | ||||  ||| ||| ||| || ||||||||
Db    275 AGGCCTATACAGCTGCACACCCCTAGGAACTTTCAGTGTCTATCAGGGAGGAGAGCTGAAGAGA 334

Qy    121 acggataccattttctggccaggttgtctcctggttaaacgctgtggtgggaactgtgcc 180
          ||| | |||| |||||| |||||||||| | ||| |||  ||||| |||||||| |||||||
Db    335 ACAGACCACCATCTTCTGGCCAGGCTGCCTCGTCGTTCAAGCGTTGTGGCGGGAACTGTGCC 394

Qy    181 tgttgtctccacaattgcaatgatgatgtcccaagcaaagttactaaaaaatac 240
          ||| | |||  ||||  |||||||||||||||||| ||  |||||||||||||
Db    395 TGCTGCATGCACACCTGCAATGAGTGCAGTGTCAGTGTCCCCAGCAAAGTCACCAAGAAATAT 454

Qy    241 cacgaggtcctcagttgagaccaaagaccggtgcaggggattgcacaaatcactcacc 300
          |||||||||||| |||||||||  |||| | |||||| ||||| |  |||||||||
Db    455 CACGAGGTCCTTCAGTTGAGACCAAGACCAGACAGAGCGGGGGCTGCACAAGTCCCTCACC 514

Qy    301 gacgtgccctggagcaccatgagagagtgtgactgtgtgc 342
```

```
Db    515 GACGTGCCCTGGAGCACCACCAGGAGGAGTGTGACTGCGTGTGC 556
          |||||||||||||||| ||||||||||||||||||||||||||||

RESULT         7
BF151355
LOCUS          BF151355       564 bp   mRNA                    EST
DEFINITION     uz15bi2.y1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3669119 5'
               similar to TR:Q9QY71 Q9QY71 FALLOTEIN. ;, mRNA sequence.
ACCESSION      BF151355
VERSION        BF151355.1  GI:11032750
KEYWORDS       EST.
SOURCE         house mouse.
ORGANISM       Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1  (bases 1 to 564)
AUTHORS        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
               Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Washington University Genome Sequencing Center
               Clone distribution: NCI-CGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               image.llnl.gov/image/html/iresources.shtml

               MGI:1429887
               Seq primer: -40RP from Gibco
               High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1..564
                     /organism="Mus musculus"
                     /strain="C57/B6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3669119"
                     /clone_lib="NCI_CGAP_Mam5"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies; Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys;
                     NIH"
BASE COUNT     157 a    130 c    140 g    137 t
ORIGIN

Query Match             72.7%;  Score 261.8;  DB 145;  Length 564;
Best Local Similarity   83.9%;  Pred. No. 3.9e-71;
Matches  296;  Conservative   0;  Mismatches  57;  Indels   0;  Gaps    0;

Qy      1 gctttgttttggaagaaaatccagagtggtggatctgaacctctaacagagaggagta 60
          ||||| | ||| |||| || ||||||||| || |||||| || || |  |||||||||||
Db    211 GCTTTCCTGTATGGGAAAAAAGCAAAGTGGTGAATCTGAATCTCCTCAAGGAGGAGGTA 270

Qy     61 agattatacagctgcacacctgtaacttctcagtgtccataaggaagaactaaagaga 120
          || |||| |||||| ||| ||||  | |||| ||| ||| ||| || ||||||||||||
Db    271 AAACTCTACAGCTGCACACCCCGGAACTTCTCAGTGTCCATACGGGAAGAGCTAAAGAGAGG 330

Qy    121 accggataccattttctggccaggttgtctcctggttaaacgctgtggtgggaactgtgcc 180
          ||||| | |||| |||||| |||||||| | ||| |||  ||||| |||||||| |||||||
Db    331 ACAGATACCATATTCTGGCCAGGTTGTTCTCCTGGTCAAGCGCTGTCAAGGGCTGTGAGGAAGTTGCC 390

Qy    181 tgttgtctccacaattgcaatgatgatgtcccaagcaaagttactaaaaaatac 240
          ||| | |||  ||||  |||||||||| ||||| || ||  |||||||||||||
Db    391 TGTTGCTCCATAATTGCAATGAATGCAATGTCAGTGTGTCCCACGTACACAAAAGTAC 450
```

Best Available Copy

```
Qy   241  cacgaggtccttcagttgagaccaaagaccggtgtcaggggattgcacaaatcactcacc  300
               |||||||||||||||  |||||||||  || ||||| ||  |||||||||  |||| |
Db   451  CATGAGGTCCTTCAGTTGAGACCAAAAACTGCATAAAGGATTGCATAAAGAGTACTCAAT  510

Qy   301  gacgtggccctggagcaccatgaggagtgtgactgtgtgtcagagggagcac  353
               || ||||| ||||||||||||||||||| ||||| |||| |||||
Db   511  GATGTGGCTGCTTGAACACCACGAGGAATGTGGCTGGTGTGTAGAGGGAACGC  563
```

RESULT 8
BF670092
LOCUS   BF670092   874 bp   mRNA   EST   21-DEC-2000
DEFINITION 602119417F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4276493 5',
            mRNA sequence.
ACCESSION   BF670092
VERSION   BF670092.1 GI:11943987
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 874)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1097 row: a column: 06
            High quality sequence stop: 598.
FEATURES   Location/Qualifiers
     source   1..874
                     /organism="Homo sapiens".
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4276493"
                     /clone_lib="NIH_MGC_56"
                     /tissue_type="primitive neuroectoderm"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggcgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT   227 a   208 c   245 g   194 t
ORIGIN

Query Match   71.1%;   Score 255.8;   DB 168;   Length 874;
Best Local Similarity 92.8%;   Pred. No. 3.3e-69;
Matches 335; Conservative 0; Mismatches 17; Indels 9; Gaps 6;

Qy   5 ttgtttttggaagaaatccagagtggtgatctgaaccttctaacagagaggtaagat 64
          ||| ||| ||||||||||||| |||||||||||||||||||||||||||||||||||
Db   69 TTGGTTTGGGAAGAAAATCCAGAGTGGTGGTCTGGATCTGAACCTTCTAACAGAGAGGTAAGAT 128

Qy   65 tata--cagctgcacacctcgtaacttctcagtgtccataa-gggaagaactaaagagaa 121
          |||| |||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   129 TATAGCATGCTGCACACCTCGTAACTTCTCAGTGTCCATAATGGGAAGAACTAAAGAGAA 188

Qy   122 ccgataccat-tttctgccaggttgtcctggttaaacgctgtggtgggaactgtgcc 180
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   189  CCGATACCATGTTTCTGGCCAGGTTGTCTCCTGGTTGTCTCCTGGTTAAACGCTGTGGTGGGAACTGTGCC  248

Qy   181  tgtgtctccacaattgcaatgaatgtcaa-tgtgtccaagcaaagttactaaaaaata  239
          ||| ||||||||||||||||||||||||  |||||||||||||||||||||||||||||
Db   249  TGTGGTCTCCACAATTGCAATGAATGTCAAGTGTCCCAAGCAAAGTTACTAAAAAATA  308

Qy   240  cca-cgagtccttcagttgagaccaaagaccggtcaggggattgcacaaatcactca  298
          ||| || |||||||||||||||||||||||||||||||||||||| || |||||||
Db   309  CCAGCGAGGGTCCTTCAGTTGAGACCAAAGACCGGTGTCAGGGGATGGCCACAAATCACTCA  368

Qy   299  ccgacgtggccctggagcaccatgaggagtgtg---actgtgtgcagaggggacacag  355
          |||||||||| |||||| |||||||||||||    |||||||||||||||||||||||
Db   369  CCGACGTGGCCCCTGGAGCACCGCCATGAGGAGTGTGACTGGTGTGTGCAGAGGGGAGGCACAA  428

Qy   356  g  356
          |
Db   429  G  429
```

RESULT 9
BG243001
LOCUS   BG243001   910 bp   mRNA   EST   13-FEB-2001
DEFINITION 602355974F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4483938 5',
            mRNA sequence.
ACCESSION   BG243001
VERSION   BG243001.1 GI:12752725
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 910)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10323 row: d column: 19
            High quality sequence stop: 690.
FEATURES   Location/Qualifiers
     source   1..910
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4483938"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor; biopsy sample"
                     /dev_stage="3 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies, Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT   242 a   220 c   250 g   198 t
ORIGIN

Query Match   69.5%;   Score 250.2;   DB 175;   Length 910;
Best Local Similarity 84.7%;   Pred. No. 1.9e-67;
Matches 304; Conservative 0; Mismatches 53; Indels 2; Gaps 2;

Qy   1 gctttgtttttggaagaaaatccagagtggtggatctgaacttctgaacctctaacagagaggta 60
          ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   155 GCTTTCCTGTATGGGAAAAAAGCAAAGTGGTGAATCTGGTGAAATCCTCAAGGAAGAGGTA 214

Qy   61 agattatacagctgcacacctcgtaacttctcagtgtccataagggaagaactaaagaga 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Query Match                 63.9%;  Score 230.2;  DB 143; Length 447;
Best Local Similarity       85.6%;  Pred. No. 2.9e-61;
Matches  256; Conservative    0; Mismatches  43; Indels   0; Gaps   0;

Qy     1 gcttttgttttggaagaaatccagagtggtggatctgaacctttcaacagaggagta  60
         |||||||  ||| |||||||||||||||||    |||| |||||||||||| |||||
Db   148 GCTTTCCTGTATGGAAAAAAAGCAAAGTGTGAATCTGAATCTCCTCAAGGAAGAAGGTA 207

Qy    61 agattatacagctgcacacctgtaacttctcagtgtccataaggaggaactaaagaga 120
         | |  |||||||||||||| ||  ||||||||||||||||  |||||||| |||||||
Db   208 AAACCTCACAGCTGCACACCCCGGACACTTCCAGTGTCCATACGGAGCTAAAGAGG    267

Qy   121 acgataccattttctggccaggtgtctcctggttaaacgctgtgggactgtgcc     180
         |||||||||| |||||||||||| |||||||||| ||| ||||||| |  |||||
Db   268 ACAGATACCATATTCTGGCCAGGTTGTCTCCTGGTCAAGCCGTGTCCCACGTAAAGTAC 327

Qy   181 tgttgtctccacaattgcaatgaatgtcccaagcaaagttactaaaaaatac        240
         |||||||||||  |||||||||||||||    |||  |||||||   ||| ||
Db   328 TGTTGTCTCCATAATTGCAATGAATGTCAGTGTCTCCACGTAAAGTTACAAAAAAGTAC 387

Qy   241 cacgaggtccctcagttgagaccaaagaccggtgtcagggattgcacaaactcactcac 299
         || |||||| ||||||| |||||||||  | ||| | |||||||||| || ||||||||
Db   388 CATGAGGTCCTTCAGTGAGACCAAAAACTGAGTCAAGGATTGCATAAAGTCACTCAC    446


RESULT 11
BG118707
LOCUS       BG118707        968 bp     mRNA            EST        30-JAN-2001
DEFINITION  602348280F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4443498 5',
            mRNA sequence.
ACCESSION   BG118707
VERSION     BG118707.1  GI:12612213
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 968)
AUTHORS     Strausberg,R.
TITLE       NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10217 row: o  column: 19
            High quality sequence start: 8
            High quality sequence stop: 574.
FEATURES             Location/Qualifiers
     source          1..968
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4443498"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT    313 a    230 c    219 g    206 t
ORIGIN


Query Match                 58.1%;  Score 209.2;  DB 174; Length 968;
Best Local Similarity       94.2%;  Pred. No. 1.3e-54;
Matches  228; Conservative    0; Mismatches  13; Indels   1; Gaps   1;
```

```
Db   215 | | | ||||||||||||| ||  |||||||||||||||| ||||||||
         AAACCTCACAGCTGCACACCCCGGACACTTCCAGTGTCCATACGGAGAGCTAAAGAGG   274

Qy   121 acgataccattttctggccaggttgtctcctggttaaacgctgtgggactgtgcc     180
         ||||||||| |||||||||||| |||||||||| ||| |||| |||  ||||||
Db   275 ACAGATACCATATTCTGGCCAGGTTGTCTCCTGGTCAAGCCGTGTCCCACGTAAAGTAC 333

Qy   181 tgttgtctccacaattgcaatgaatgtcccaagcaaagttactaaaaaatac        240
         ||||||||||  |||||||||||||||    |||  ||||||| || ||| ||
Db   334 TGTTGTCTCCATAATTGCAATGAATGTCAGTGTCTCCACGTAAAGTTACAAAAAAGTAC 393

Qy   241 cacgaggtccctcagttgagaccaaagaccggtgtcagggattgcacaaatcactcacc 300
         || |||||| ||||||| ||||||||||   || |  ||| |||||| ||  ||||||
Db   394 CATGAGGTCCTTCAGTGAGACCAAAA-AACTGAGTCAAGGATTGCATAAGTCACTCACT 452

Qy   301 gacgtggcctggagcaccatgagaggatgtgactgtgtgtgtgcagaggagcacaggagg 359
         ||| |||||| | ||||| |||||| | ||| ||| |||||||||||  |||||||||||
Db   453 GATGTGGCTCTGGAACACCACCAGGAATGTGACTGTGTGTGTGTAGAGGAAACGCAGGAGG 511


RESULT 10
BF011835
LOCUS       BF011835        447 bp     mRNA            EST        06-OCT-2000
DEFINITION  us37610.y1 Soares_NMEBA_branchial_arch Mus musculus cDNA clone
            IMAGE:3169267 5' similar to TR:Q9QY71 Q9QY71 FALLOTEIN. ;, mRNA
            sequence.
ACCESSION   BF011835
VERSION     BF011835.1  GI:10712110
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 447)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:1064727
            Seq primer: -40RP from Gibco
            High quality sequence stop: 402.
FEATURES             Location/Qualifiers
     source          1..447
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3169267"
                     /clone_lib="Soares_NMEBA_branchial_arch"
                     /tissue_type="branchial arches"
                     /dev_stage="embryo, 10.5 dpc"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAGTGGGAGCGGCGGCGCCGCATTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI site of the modified pT7T3 vector. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT    126 a    104 c    108 g    109 t
ORIGIN
```

Wed Jul 18 08:51:37 2001   us-09-457-066-1_copy_829_1188.rst

Db   13 CGCTGTGGAGGAAATTGTGCCTGTTGTGTCTCCATAATTGCAATGAATGTCAGTGTGTCCCA 72
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   220 agcaaagttactaaaaataccacgaggtccttcagttgacaccaaagaccggtgtcagg 279
                |||||||||||||||||| ||||||||||||||||||||||||||||||||| ||||
Db   73 CGTAAAGTTACCAAAAAAGTACCATGAGGTCCTTCAGTTGACACCAAAAACTGGAGTCAAG 132

Qy   280 ggattgcacaaatcactcaccgacgtggccctggagcaccatgagagagtgtgactgtgtg 339
                ||||||| ||||| |||||| ||| | |||| ||||||||||||||||||||||||
Db   133 GGATTGCATAAGTCACTCACTCAGTGCGGCTCTGGAACACCACCAGGAATGTGACTGTGTG 192

Qy   340 tgcagaggagcacaggagg 359
                |||||||||||||||||||
Db   193 TGTAGAGGAAACGCAGGAGG 212

RESULT 13
AA759138/c
LOCUS   AA759138   425 bp   mRNA   EST   29-DEC-1998
DEFINITION ah78f03.s1 Soares_testis_NHT Homo sapiens cDNA clone 1321757 3',
            mRNA sequence.
ACCESSION   AA759138
VERSION   AA759138.1 GI:2807001
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 425)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1417   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 385.
FEATURES   Location/Qualifiers
     source   1..425
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="1321757"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Cloptech Laboratories
                     , Inc. and primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAGTGGAGCGGCCGCCCAATTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT   100 a   108 c   87 g   130 t
ORIGIN

Query Match   39.3%; Score 141.4; DB 11; Length 425;
Best Local Similarity 98.7%; Pred. No. 1.8e-33;
Matches 153; Conservative   0; Mismatches   1; Indels   1; Gaps   1;

Qy   206 gtcaatgtgtcccagcaagcaaagttactaaaaataccacgaggtccttcagttgagaccaa 265
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   425 GTCAATGTGTCCCAGCAAGCAAAGTTACTAAAAATACCACCAGGTCCTTCAGTTGAGACCAA 366

Qy   119 gaaccgataccatttctggccaggttgtctcctggttaaacgctgtggtgggaactgtg 178
                || ||| || ||||||||||||| ||| |||||| ||||||||||||||||||||
Db   254 GATCTGAACCTTCTAACAGAGGAGGTTGTCTCCTGGTTAAACGCTGTGGTGGGAACTGTG 313

Qy   179 cctgtgtcccacacaattgcatgaatgtcaatgtgtcccagcaagtgcaaagttactaaaaaaat 238
                ||||||||||||||| ||| |||||||||||||||||||||||||||| ||||||||||||
Db   314 CCTGTTGTCCCACAATTGCAATGAATCAATGTGTCCCAGCAAAGTTACTAAAAAAT 373

Qy   239 accacgaggtccttcagttgagaccaaaagaccggtgtcagttgcacaaatcactca 298
                |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   374 ACCACGAGGTCCTTCAGTTGAGACCAACAGACCGGTCAGGGGATTGCACAAATCACTCA 433

Qy   299 ccgacgtgccctggagcaccatgagagagtgtgtgactgtgtgcagaggagcacaggag 358
                ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   434 CCGACGTGG-CCTCGGAGCACCATGAGGAGTGTGACTGTGTGCAGAGGGAGCACAGGAG 492

Qy   359 ga 360
                ||
Db   493 GA 494

RESULT 12
BF163629
LOCUS   BF163629   902 bp   mRNA   EST   30-OCT-2000
DEFINITION 601769732F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3989215 5',
            mRNA sequence.
ACCESSION   BF163629
VERSION   BF163629.1 GI:11044002
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 902)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9198 row: o column: 08
            High quality sequence stop: 743.
FEATURES   Location/Qualifiers
     source   1..902
                     /organism="Mus musculus"
                     /strain="C2ECH II (feral)"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3989215"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT   248 a   203 c   241 g   210 t
ORIGIN

Query Match   44.0%; Score 158.4; DB 145; Length 902;
Best Local Similarity 87.0%; Pred. No. 1e-38;
Matches 174; Conservative   0; Mismatches 26; Indels   0; Gaps   0;

Qy   160 cgctgtggtgggaactgtgcctgtgtcctgttgtctccacaattgcaattgaatgtcaatgtgtcccca 219
                |||||| || || ||||||| || || ||||||||||||||| ||| ||||||||||||

Qy   266 agaccggtcgtgtggggattgcaaatcactcaccgacgtggccctgagcaccatgagg 325
           ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   365 AGACCGGTGTCAGGGGA-TGCACAAATCACTCACCGACGTGGCCCTGGAGCACCATGAGG 307

Qy   326 agtgtgactgtgtgcagagggagcacaggagga 360
           |||||||||||||||||||||||||||| |||
Db   306 AGTGTGACTGTGTGTGCAGAGGGAGCACAGGGGGA 272

RESULT 14
BE374398
LOCUS   BE374398   521 bp   mRNA   EST   21-JUL-2000
DEFINITION   601227568F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3585833 5',
                mRNA sequence.
ACCESSION   BE374398
VERSION   BE374398.1   GI:9319761
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 521)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: Gilbert Smith, Ph.D.
                cDNA Library Preparation: Life Technologies, Inc.
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLAM8746 row: g   column: 18
                High quality sequence stop: 521.
FEATURES   Location/Qualifiers
     source   1..521
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3585833"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="3 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT   150 a   128 c   126 g   117 t
ORIGIN

Query Match   38.8%;   Score 139.6; DB 166; Length 521;
Best Local Similarity 86.8%; Pred. No. 7e-33;
Matches 165; Conservative   0; Mismatches 24; Indels 1; Gaps 1;

Qy   170 ggaactgtgctgttctccacaattgcaatgaatgtcatgtgtcccaagcaaagtta 229
           |||| ||||| |||||||||| |||||||||||||||||||||||||||||| |||||
Db   4 GGAAATGTGCCTG-TGTCTCCATAATTGCAATGAATGTCAGTGTCCCACGTAAAGTTA 62

Qy   230 ctaaaaataccacgaggtccttcagttgagaccaaagaccggtgtcagggggattgcaca 289
           | ||||| |||| ||||||||||||||||||||| || |||||||||||||| ||||
Db   63 CAAAAAAGTACCATGAGGTCCTTCAGTTGAGACCAAAAACTGGAGTCAAGGGATTGCATA 122

Qy   290 aatcactcaccgacgtggccctggaccatgagaggtgtgactgtgtgtcagagagga 349
           |||||||||||||||| |||||| |||||||||||||| ||||||||||||||||||
Db   160 GCTTTCCTGRATGTGAAAAAAAGCAAAGTGGTGAATCTGAATCTCCTCAAGGAAGAGGTA 219

Qy   350 gcacaggagg 359
           | ||||||||
Db   220 AAACTTACACAGCTGCCACACCCCGGAACTTCTCAGTGTCCATACGTGGGAAGAGGAGTA... (cut)

Db   183 ACGCAGGAGG 192

RESULT 15
AI020581
LOCUS   AI020581   324 bp   mRNA   EST   16-JUN-1998
DEFINITION   ua96c08.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
                IMAGE:1365326 5', mRNA sequence.
ACCESSION   AI020581
VERSION   AI020581.1   GI:3234917
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 324)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
                Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
                Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
                Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
                Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
                WashU-HHMI Mouse EST Project
                Washington University School of MedicineP
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: mouseest@watson.wustl.edu
                This clone is available royalty-free through LLNL; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                MGI:898546
                Seq primer: -28m13 rev2 ET from Amersham
                High quality sequence stop: 283.
FEATURES   Location/Qualifiers
     source   1..324
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1365326"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with NotI and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT   82 a   81 c   81 g   80 t
ORIGIN

Query Match   32.9%;   Score 118.6; DB 15; Length 324;
Best Local Similarity 82.4%; Pred. No. 2.3e-26;
Matches 136; Conservative   0; Mismatches 29; Indels 0; Gaps 0;

Qy   1 gctttgttttggaagaaatccagagtggtggatctgaacctctaacagagaggagta 60
           |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   160 GCTTTCCTGRATGTGAAAAAAAGCAAAGTGGTGAATCTGAATCTCCTCAAGGAAGAGGTA 219

Qy   61 agattatacagctgcacacctcgtaacttctcagtgtccataaggaagaactaaagaga 120
           | || |||||||||||||||||| || ||||||||||||||||||||| ||||||||
Db   220 AAACTTACACAGCTGCCACACCCCGGAACTTCTCAGTGTCCATACGTGGGAAGAGGAGTA 279

Wed Jul 18 08:51:37 2001

us-09-457-066-1_copy_829_1188.rst

```
Qy   121 accgataccatttctgccaggttgtctcctggttaaacgctgt 165
         || ||||||||| |||||||||||||||||| || ||||||
Db   280 ACAGATACCATATTCTGGCCAGGTTGTCTCCTGGTCAAGCGCTGT 324
```

Search completed: July 18, 2001, 03:49:55
Job time: 13553 sec